

AMENDMENT

Please amend the application as follows:

In the specification:

Please replace the title with the new title:

-- ALPHA GALACTOSIDASES AND METHODS FOR MAKING AND USING
THEM --

Please insert the following paragraph before the paragraph beginning at page 1, line 28:

a¹ -- *Thermococcus alcaliphilus* AEDIII2RA α -galactosidase 18GC has been deposited
with the ATCC located at 10801 University Blvd., Manassas, VA 20110-2209, on September
10, 2002. The Patent Deposit Designation is PTA-4654. --

Please replace the paragraph beginning at page 24, line 24, with the following:

a² -- As representative examples of expression vectors which may be used, there may be
mentioned viral particles, baculovirus, phage, plasmids, phagemids, cosmids, fosmids, bacterial
artificial chromosomes, viral DNA (e.g., vaccinia, adenovirus, fowl pox virus, pseudorabies and
derivatives of SV40), P1-based artificial chromosomes, yeast plasmids, yeast artificial
chromosomes, and any other vectors specific for specific hosts of interest (such as bacillus,
aspergillus and yeast). Thus, for example, the DNA may be included in any one of a variety of
expression vectors for expressing a polypeptide. Such vectors include chromosomal,
nonchromosomal and synthetic DNA sequences. Large numbers of suitable vectors are known
to those of skill in the art, and are commercially available. The following vectors are provided
by way of example; Bacterial: pQE vectors (Qiagen), pBLUESCRIPT® plasmids, pNH vectors,
(lambda-ZAP vectors (Stratagene); ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia);
Eukaryotic: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, pSVLSV40 (Pharmacia).
However, any other plasmid or other vector may be used so long as they are replicable and viable
in the host. Low copy number or high copy number vectors may be employed with the present
invention. --

Please replace the paragraph beginning at page 43, line 30, with the following:

a³ -- Particular bacterial vectors which may be used include the commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017), pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden), GEM1 (Promega Biotec, Madison, Wis., USA) pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174 pBLUESCRIPT® II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene), ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia), pKK232-8 and pCM7. Particular eukaryotic vectors include pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). However, any other vector may be used as long as it is replicable and viable in the host cell. --

Please replace the paragraph beginning at page 58, line 29, with the following:

a⁴ -- A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequence for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482, 1981, by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443, 1970, by the search for similarity method of person & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444, 1988, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection. Other algorithms for determining homology or identity include, for example, in addition to a BLAST™ program (Basic Local Alignment Search Tool at the National Center for Biological Information), ALIGN, AMAS™ (Analysis of Multiply Aligned Sequences), AMPS™ (Protein Multiple Sequence Alignment), ASSET™ (Aligned Segment Statistical Evaluation Tool), BANDS, BESTSCOR, BIOSCAN™ (Biological Sequence Comparative Analysis Node), BLIMPS (BLOCKS IMPROVED Searcher), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W, CONSENSUS™, LCONSENSUS, WCONSENSUS, Smith-Waterman

a4
algorithm, DARWIN™, Las Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framealign, FRAMESEARCH™, DYNAMIC™, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment Program), GENAL, GIBBS™, GENQUEST™, ISSCT™ (Sensitive Sequence Comparison), LALIGN (Local Sequence Alignment), LCPT™ (Local Content Program), MACAW™ (Multiple Alignment Construction & Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN, PIMA (Pattern-Induced Multi-sequence Alignment), SAGA™ (Sequence Alignment by Genetic Algorithm) and WHAT-IF. Such alignment programs can also be used to screen genome databases to identify polynucleotide sequences having substantially identical sequences. A number of genome databases are available, for example, a substantial portion of the human genome is available as part of the Human Genome Sequencing Project (J. Roach) (Gibbs, 1995). At least twenty-one other genomes have already been sequenced, including, for example, *M. genitalium* (Fraser *et al.*, 1995), *M. jannaschii* (Bult *et al.*, 1996), *H. influenzae* (Fleischmann *et al.*, 1995), *E. coli* (Blattner *et al.*, 1997), and yeast (*S. cerevisiae*) (Mewes *et al.*, 1997), and *D. melanogaster* (Adams *et al.*, 2000). Significant progress has also been made in sequencing the genomes of model organism, such as mouse, *C. elegans*, and *Arabidopsis sp.* Several databases containing genomic information annotated with some functional information are maintained by different organization, and are accessible via the internet, for example, the website for The Institute For Genomic Research; the genetics website for the University of Wisconsin - Madison; the Stanford University Genomic Resources website; the website for the HIV database; the website for the National Center for Biotechnology Information; the website for the European Bioinformatics Institute; the website for the Institut Pasteur; and the website for the Whitehead Institute/MIT Center for Genome Research.

--
Please replace the paragraph beginning at page 60, line 7, with the following:

a5
-- One example of a useful algorithm is BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*, Nuc. Acids Res. 25:3389-3402, 1997, and Altschul *et al.*, J. Mol. Biol. 215:403-410, 1990, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website. This algorithm involves first

a⁵ identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectations (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, 1989) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands. --

Please replace the paragraph beginning at page 61, line 21, with the following:

-- The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (*i.e.*, aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet *et al.*, Science 256:1443-1445, 1992; Henikoff and Henikoff, Proteins 17:49-61, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, *e.g.*, Schwartz and Dayhoff, eds., 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation). BLAST programs are accessible through the U.S. National

a⁶

Library of Medicine, e.g., at the website for the National Center for Biotechnology Information.

Please replace the paragraph beginning at page 66, line 19, with the following:

a7
-- Figure 5 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group. Alternatively, the features may be structural polypeptide motifs such as alpha helices, beta sheets, or functional polypeptide motifs such as enzymatic active sites, helix-turn-helix motifs or other motifs known to those skilled in the art. --

Please replace the paragraph beginning at page 71, line 6, with the following:

a8
-- Colonies containing pBLUESCRIPT® plasmids with random inserts from the organism *Thermococcus alcaliphilus* AEDII12RA were obtained from an original XZAP2 genomic library generated according to the manufacturer's (Stratagene) protocol. The clones were then excised from λZAP2 to pBLUESCRIPT®. The clones were excised to pBLUESCRIPT® according to the method of Hay and Short. (Hay, B. and Short, J. Strategies, 1992, 5:16.) The resulting colonies were picked with sterile toothpicks and used to singly inoculate each of the wells of 96-well microtiter plates. The wells contained 250 μL of LB media with 100 μg/ml methicillin, and 10% v/v glycerol (LB Amp/Meth, glycerol). The cells were grown overnight at 37°C. without shaking. This constituted generation of the "Source GeneBank"; each well of the Source GeneBank thus contained a stock culture of *E. coli* cells, each of which contained a pBLUESCRIPT® plasmid with a unique DNA insert. --
